JOURNAL OF MULTIVARIATE ANALYSIS 20, 244–250 (1986)

The Distribution of a Generalized Least Squares Estimator with Covariance Adjustment

M. G. KENWARD

The Mathematical Institute, The University, Canterbury, Kent, CT2 7NF, United Kingdom

Communicated by C. G. Khatri

The distribution theory is developed for a generalized least squares estimator of the growth curve model. A special case of the estimator is the maximum likelihood estimator which is weighted by the sample covariance matrix. The distribution of two conditional forms of the estimator are derived and from these its density is obtained. Two general pivots and their distributions are derived from the conditional forms and special cases of these are investigated. The results obtained are linked to earlier work. © 1986 Academic Press, Inc.

1. INTRODUCTION

In this paper the distribution of a class of generalized least squares estimators of the parameters of the growth curve model is investigated. The model was introduced in [10] and can be summarized as follows for the $(n \times q)$ matrix or random variables Y,

$$Y \sim N(A\Gamma T; I_n; \Sigma), \tag{1.1}$$

where $A(n \times a, n > a,$ rank $a)$ is a known design matrix, $T(t \times q, t < q,$ rank $t)$ is a known matrix of regression variables, and $\Gamma(a \times t)$ is a matrix of unknown parameters. The maximum likelihood estimator of Γ is

$$\hat{\Gamma}_L = (A'A)^{-1} A'YS^{-1}T'(TS^{-1}T')^{-1},$$

where $S = Y'\{I_n - A(A'A)^{-1}A'\}Y$ and it is assumed that $n - a > q$ [8]. This can be re-written

$$\hat{\Gamma}_L = (A'A)^{-1} A'Y_{T \cdot V},$$

Received October 15, 1982; revised January 24, 1986.

AMS 1980 subject classifications: Primary 62H10.

Key words and phrases: analysis of covariance, conditional model, generalized least squares, growth curve model, matrix-variate distribution.

0047-259X/86 $3.00

where $Y_{T \cdot V} = Y_T - YV(V'SV)^{-1} V'ST'(TT')^{-1}$ and $Y_T = YT'(TT')^{-1}$ for $V\{q \times (q-t)\}$, satisfying $TV = 0$ and $V'V = I_{q-t}$. We see from this representation that $\hat{\Gamma}_L$ can be regarded as a covariance adjusted estimator in which the transformed variables YV, which have zero expectation under model (1.1) are used to adjust the simple least squares estimator $(A'A)^{-1} A'Y_T$. It follows that it may be desirable to use only a subset of these covariates, see, for example [11, Sect. 5; 12, Sect. 3]. If the columns of $W(q \times w)$ are a selection of the columns of V, then a more general form of the estimator is given by

$$\hat{\Gamma} = (A'A)^{-1} A'Y_{T \cdot W}, \tag{1.2}$$

where $Y_{T \cdot W}$ is obtained by replacing V by W in the definition of $Y_{T \cdot V}$. Note that $\hat{\Gamma}$ is defined when $n - a > t + w$ and this will be assumed from now on.

Although the two approaches, one via generalized least squares and the other via covariance adjustment, produce the same estimator, each provides a different viewpoint. It will be seen below how these suggest two complementary expressions for the distribution of $\hat{\Gamma}$. This is investigated for the more general form $L\hat{\Gamma}M$ in Section 2. These expressions are used in the derivation of two pivots in Section 3. Some special cases of these pivots are given and related to statistics which are more familiar in practice. These statistics are presented pragmatically and no attempt is made to derive optimal test criteria. Such an approach can be found in [7], where a review is given of this aspect of the subject.

2. Distribution of the Estimator

In order to simplify the theory, $\hat{\Gamma}$ defined in (1.2) will be standardized using a transformation that is related to the canonical representation in [4, 5]. For more generality the estimator $L\hat{\Gamma}M$ of $L\Gamma M$ will be considered, where $L(l \times a,\ l \leqslant a)$ and $M(t \times m,\ m \leqslant t)$ are fixed matrices of full rank. It will be shown that $L\hat{\Gamma}M$ can be expressed as

$$L\hat{\Gamma}M = L\Gamma M + C^{-1}QR^{-1}, \tag{2.1}$$

where C, R, and Q are defined as follows:

(i) C ($l \times l$ non-singular) satisfies $(C'C)^{-1} = L(A'A)^{-1}L'$;

(ii) $R(m \times m)$ is the unique lower-triangular matrix with positive diagonal elements satisfying, say,

$$(RR')^{-1} = M'(RR')^{-1}T\{\Sigma - \Sigma W(W'\Sigma W)^{-1} W'\Sigma\}\, T'(TT')^{-1}M = \Phi;$$

(iii) $Q = X_1 - X_2 U_{22}^{-1} U_{21}$, where $X_1(l \times m)$, $X_2(l \times w)$, $U_{22}(w \times w)$, and $U_{21}(w \times m)$ are distributed according to

$$[X_1, X_2] = X \sim N(0; I_l; I_p),$$

$$\begin{bmatrix} U_{11}, & U_{12} \\ U_{21}, & U_{22} \end{bmatrix} = U \sim W_p(n-a; I_p), \qquad p = m + w,$$

with X and U independent.

To derive the equivalence in (2.1) define $X = [Z_T - CL\Gamma, Z_W]\, P$ and $U = P'DP$, where $Z_T = CL(A'A)^{-1} A'Y_T M$, $Z_W = CL(A'A)^{-1} A'YW$, $D = T'_* S T_*$, $T_* = [T'(TT')^{-1} M, W]$, and P is the unique lower-triangular matrix with positive diagonal elements satisfying $P'T'_* \Sigma T_* P = I_p$. It follows that X and U are independent with the distributions given above. Partitioning D and P like U gives

$$X_1 = (Z_T - CL\Gamma M)\, P_{11} + Z_W P_{21}$$

$$X_2 U_{22}^{-1} U_{21} = Z_W D_{22}^{-1} D_{21} P_{11} + Z_W P_{21}$$

and

$$L\hat{\Gamma} M = C^{-1}(Z_T - Z_W D_{22}^{-1} D_{21}).$$

Thus

$$\begin{aligned} Q &= X_1 - X_2 U_{22}^{-1} U_{21} \\ &= CL(\hat{\Gamma} - \Gamma)\, MP_{11}. \end{aligned} \tag{2.2}$$

Note also that $P_{11} = R$. This follows from the equality of PP' and $(T'_* \Sigma T_*)^{-1}$, which implies that $P_{11} P'_{11} = \Phi^{-1}$. Setting $R = P_{11}$ in (2.2) and rearranging gives (2.1).

The distribution of Q will now be investigated using two complementary conditional expressions. The first is of the conditional distribution of Q given U. Since $Q = X_1 - X_2 U_{22}^{-1} U_{21}$ this can be written

$$Q \mid H \sim N(0; I_l; I_m + H), \tag{2.3}$$

where $H = U'_* U_{22}^{-1} U_*$ for $U_* = U_{22}^{-1/2} U_{21}$. From the Bartlett decomposition theorem U_* and U_{22} are independent and $U_* \sim N(0; I_w; I_m)$, thus $H \sim F(w; n-a-w+m; I_m)$ [9, Theorem 4.1]. The distribution of H is singular unless $w \geqslant m$. Together with (2.3) this gives a complete description of the distribution of Q. The second expression is obtained by working conditionally on the value of the concomitant variables YW. However the same expression can be obtained directly from (2.3) by using th results in [2, Sect. 5]. The sphericity of the standard matrix-variate normal distribution, the rotatability of the matrix-variate F distribution, together with the extendibility of both, imply that

$$\begin{aligned} Q \mid G &\sim N(O; I_l + G; I_m), \\ G &\sim F(w, n-a-w+l; I_l), \end{aligned} \tag{2.4}$$

where $G = X_2 U_{22}^{-1} X_2'$ and has a singular distribution unless $w \geqslant l$. The two corresponding conditional expressions for $L\hat{\Gamma}M$ are

$$L\hat{\Gamma}M \mid H \sim N(L\Gamma M; L(A'A)^{-1}L; (R')^{-1}(I_m + H)R^{-1}),$$
$$L\hat{\Gamma}M \mid G \sim N(L\Gamma M; C^{-1}(I_l + G)(C')^{-1}; \Phi).$$

When l or m is equal to 1 these take a particularly simple form. For example, when $l = 1$,

$$\begin{aligned} L\hat{\Gamma}M \mid u &\sim N(L\Gamma M; u^{-1}L(A'A)^{-1}L'\Phi) \\ u &\sim B\{(n-a-w+1)/2; w/2\}, \end{aligned}$$

where $u = (1+G)^{-1}$ and similarly when $m = 1$.

In order to derive the marginal densities of Q and $L\hat{\Gamma}M$ we rewrite (2.4) as

$$\begin{aligned} Q \mid Z &\sim N(O; I_l + ZZ'; I_m), \\ Z &\sim T(n-a-w+l; I_l : I_w), \end{aligned}$$

since the density of Z is defined whenever that of $\hat{\Gamma}$ is, whereas that of G is defined only when $l < w$. Note that we could equally well work with the other conditional expression (2.3). The joint density of Q and Z can be written

$$\begin{aligned} &(2\pi)^{-lm/2} |I_l + ZZ'|^{-m/2} \exp\{-\tfrac{1}{2} \operatorname{tr} QQ'(I_l + ZZ')^{-1}\} \\ &\cdot K_1 |I_l + ZZ'|^{-(n-a+l)/2} \end{aligned}$$

where $K_1 = \pi^{-lw/2} \Gamma_l\{(n-a+l)/2\}/\Gamma_l\{(n-a-w+l)/2\}$. Integrating Z, we get the marginal density of Q:

$$K_1 K_2 (2\pi)^{-lm/2} \exp(-\tfrac{1}{2} \operatorname{tr} QQ')\, {}_1F_1(w/2; (n-a+l+m)/2; \tfrac{1}{2}QQ'),$$

for $K_2 = \pi^{lw/2} \Gamma_l\{(n-a-w+l+m)/2\}/\Gamma_l\{(n-a+l+m)/2\}$. Transforming to $L\hat{\Gamma}M = L\Gamma M + C^{-1}QR^{-1}$ we get the marginal density of $L\hat{\Gamma}M$:

$$\begin{aligned} &K_1 K_2 (2\pi)^{-lm/2} |\Phi|^{-l/2} |L(A'A)^{-1}L'|^{-m/2} \exp(-\tfrac{1}{2} \operatorname{tr} B) \\ &\cdot {}_1F_1\{w/2; (n-a+l+m)/2; \tfrac{1}{2}B\}, \end{aligned}$$

where $B = CL(\hat{\Gamma} - \Gamma) M\Phi^{-1}M'(\hat{\Gamma} - \Gamma) L'C'$.

The asymptotic distribution of $L\hat{\Gamma}M$ as $n-a\to\infty$ is normal. This can be deduced using the limiting distribution of S in the definition of $\hat{\Gamma}$, or from the characteristic function of Q, which can be written

$$\begin{aligned}\phi(T) &= E_Q\{\exp(i \operatorname{tr} TQ)\} = E_Z[E_{Q|Z}\{i \operatorname{tr} TQ\}] \\ &= E_Z[\exp\{-\tfrac{1}{2}\operatorname{tr} T'T(I_l+ZZ')\}] \\ &= \exp\{-\tfrac{1}{2}\operatorname{tr} T'T\}\, E_Z[\exp\{-\tfrac{1}{2}wT'TZZ'\}].\end{aligned}$$

Since the second term on the right tends to one as $n-a\to\infty$, we have $\phi(T)\to\exp\{-\frac{1}{2}\operatorname{tr} T'T\}$.

The normal approximation can be improved for smaller values of $n-a$ if I_l+G is replaced by its expected value $(n-a-1)\,I_l/(n-a-w-1)$, in which case the first and second order moments of the approximating normal and exact distribution are equal.

3. Two Pivots Based on the Estimator

The problem of making inferences about $L\Gamma M$ using $L\hat{\Gamma}M$ can be approached in many ways. Two relatively simple, general pivots will be considered here. Statistics based on special cases of these have been suggested in the past and some examples of these will be given below. Both pivots involve $\hat{\Phi}$, the sample estimate of Φ, defined by

$$\hat{\Phi}=(n-a-w)^{-1}M'(TT')^{-1}T\{S-SW(W'SW)^{-1}W'S\}\,T'(TT')^{-1}M.$$

Note that $(n-a-w)\hat{\Phi}\sim W_m(n-a-w;\Phi)$ and is independent of H in (2.3) and G in (2.4).

The first pivot corresponds to multivariate analysis of covariance and is derived from the conditional distribution of $L\hat{\Gamma}M$ given G (2.4). It is defined as the $(l\times l)$ matrix P_1, where

$$\begin{aligned}P_1 &= (n-a-w)^{-1}(I_l+G)^{-1/2}CL(\hat{\Gamma}-\Gamma) \\ &\quad\times M\hat{\Phi}^{-1}M'(\hat{\Gamma}-\Gamma)'L'C'(I_l+G)^{-1/2}.\end{aligned}$$

Since $(I+G)^{-1/2}CL(\hat{\Gamma}-\Gamma)M \mid G\sim N(O;I_l;\Phi)$ it follows from [9, Theorem 4.1] that $P_1\sim F(m,n-a-w-m+l;I_l)$. The non-zero eigenvalues of P_1 have important invariance properties and any "reasonable" statistic will be based on these, as in MANOVA, see, for example, [7].

The second pivot is defined as follows:

$$\begin{aligned}P_2 &= (I_l+G)^{1/2}P_1(I_l+G)^{1/2} \\ &= (n-a-w)^{-1}CL(\hat{\Gamma}-\Gamma)\,M\hat{\Phi}^{-1}M'(\hat{\Gamma}-\Gamma)'L'C'.\end{aligned}$$

From (2.4), $CL(\hat{\Gamma}-\Gamma)M \mid G \sim N(O; I_l+G; \Phi)$. Thus, $P_2 \mid G \sim F(m, n-a-w-m+l; I_l+G)$. Clearly, the distribution of P_2 is free of nuisance parameters. In order to derive the density of P_2 we shall again work with related T variables, thus avoiding problems with singular matrix variables. We can write $P_2 \sim T_2 T_2'$, where $T_2=(n-a-w)^{-1/2} Q R^{-1} \hat{\Phi}^{-1/2}$,

$$T_2 \mid Z \sim T(n-a-w-m+l; I_l+ZZ'; I_m),$$

and $Z \sim T(n-a-w+l; I_l; I_w)$ was defined in Section 2. The joint density of T_2 and Z is

$$K_1 |I_l+ZZ'|^{-(n-a+l)/2} K_3 |I_l+ZZ'|^{-(n-a-w-m+l)/2}$$
$$\times |I_l+ZZ'+T_2 T_2'|^{-(n-a-w+l)/2}$$

where $K_3=\pi^{-lm/2}\Gamma_l\{(n-a-w+l)/2\}/\Gamma_l\{(n-a-w-m+l)/2\}$ and K_1 was defined in the previous section. Integrating Z out of this joint density, we can get the following expression for the marginal density of T_2:

$$\frac{K_1 K_3}{K_4} |I+T_2 T_2'|^{-(n-a-w+l)/2}$$
$$\cdot {}_2F_1\{w/2, (n-a-w+l)/2; (n-a+m+l)/2; I_l-(I_l+T_2 T_2')^{-1}\},$$

where $K_4=\pi^{-lm/2}\Gamma_l\{(n-a+m+l)/2\}/\Gamma_l\{(n-a-w+m+l)/2\}$. As with P_1 one would expect statistics to be based on the eigenvalues of $P_2=T_2 T_2'$. It is clear from the definition of T_2 in terms of Q, that the asymptotic distribution of T_2 as $n-a \to \infty$ is $T(n-a-w-m+l; I_l; I_m)$, which in turn tends to $N(O; I_l; I_m)$, i.e., P_1 and P_2 have the same distribution in the limit. This second pivot has not been given before, to the author's knowledge, special cases which have been obtained previously are mentioned below.

Consider the case $m=1$. This includes the analysis of individual columns of $L\hat{\Gamma}$. The use of P_1 to test the null hypothesis $L\hat{\Gamma}M=0$ corresponds to the standard analysis of covariance and the null distribution of $(n-a-w)\, l^{-1} \operatorname{tr} P_1$ is $F_{l,n-a-w}$. The use of P_2 to test the same hypothesis is closely related to the "approximate" analysis of covariance described in [3, p. 257; 13; and 1]. If the hypothesis mean square $l^{-1}M'\hat{\Gamma}L'\{L(A'A)^{-1}L'\}^{-1}L\hat{\Gamma}M$ is divided by the residual mean square $(n-a-1)\, M'\hat{\Phi}M/(n-a-w-1)$, which has the same expectation under the null hypothesis, we get the alternative variance ratio with density function

$$f(x)=K_5 x^{l/2-1}(1+rx)^{-(n-a-w+l)/2}$$
$$\times {}_2F_1\{w/2, (n-a-w+l)/2; (n-a+1+l)/2; rx/(1+rx)\}, \qquad (3.1)$$

where $K_5 = r^{l/2} B\{w/2, (n-a-w+l+1)/2\}[B\{l/2, (n-a-w)/2\} B\{w/2, (n-a-w+l)/2\}]^{-1}$ and $r = l(n-a-1)/\{(n-a-w-1)(n-a-w)\}$. This is derived from the distribution of tr $T_2 T_2'$ assuming $m=1$. A selection of the upper 5% and 1% points of this distribution have been derived by the author. However, the $F_{l,n-a-w}$ distribution provides an adequate approximation for most situations.

For the special case $l=1$, which includes rows of $\hat{\Gamma}M$, both pivots and their distributions have been derived by Rao [11, 12, and references given there]. See also [6] for a comparison of confidence regions based on the two pivots. In this case $(n-a-w-m+1)m^{-1}P_1$ is a Hotelling T^2 statistic with an $F_{m,n-a-w-m+1}$ distribution and the second pivot has a null hypothesis distribution of a similar form to (3.1) above.

Acknowledgments

Financial support for some of the work reported here was provided by a U.K. SERC CASE Studentship and by a Reading University Wilkie Calvert Studentship. The author thanks Dr. W. J. Krzanowski for valuable comments on earlier drafts of this paper and a referee for suggesting several improvements.

References

[1] Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics* **13** 261–281.

[2] Dawid, A. P. (1978). Extendibility of spherical matrix distributions. *J. Multivariate Anal.* **8** 559–566.

[3] Fisher, R. A. (1932). *Statistical Methods for Research Workers.* 4th ed., Oliver & Boyd, Edinburgh.

[4] Gleser, L., and Olkin, I. (1966). A k-sample regression model with covariance. In *Multivariate Analysis* (P. R. Krishnaiah, Ed.), pp. 59–72. Academic Press, New York.

[5] Gleser, L., and Olkin, I. (1970). Linear models in multivariate analysis. In *Essays in Probability and Statistics* (R. C. Bose *et al.*, Eds.), pp. 267–292. Univ. of North Carolina Press, Chapel Hill.

[6] Gleser, L., and Olkin, I. (1972). Estimation for a regression model with an unknown covariance matrix. In *Proceedings, Sixth Berkely Sympos.*, pp. 451–568. Univ. of California Press, Berkely.

[7] Kariya, T. (1978). The general MANOVA problem. *Ann. Statist.* **6** 200–214.

[8] Khatri, C. G. (1966). A note on a MANOVA model applied to problems in growth curves. *Ann. Inst. Statist. Math.* **18** 75–86.

[9] Olkin, I., and Rubin, H. (1964). Multivariate beta distributions and independence properties of the Wishart distribution. *Ann. Math. Statist.* **35** 261–269.

[10] Pothoff, R. F., and Roy, S. N. (1964). A generalized multivariate analysis of variance model, useful especially for growth curve problems. *Biometrika* **53** 313–326.

[11] Rao, C. R. (1966). Covariance adjustment and related topics in multivariate analysis. In *Multivariate Analysis* (P. R. Krishnaiah, Ed.), pp. 87–108. Academic Press, New York.

[12] Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proceedings, Fifth Berkely Sympos.*, pp. 355–372. Univ. of California Press, Berkely.

[13] Yates, F. (1934). A complex pig-feeding experiment. *J. Agri. Sci.* **24** 511–531.